313437

(3)

(2)

DEFENCE RESEARCH AGENCY MALVERN

MEMORANDUM No. 4599

A SEGMENTAL HIDDEN MARKOV MODEL FOR SPEECH PATTERN PROCESSING

Author: M J Russell

DEFENCE RESEARCH AGENCY,
MALVERN,
WORCS.

MEMORANDUM No. 4599

Defence Research Agency,
Malvern
Memorandum 4599

A Segmental Hidden Markov Model for Speech Pattern Processing

Martin J Russell
Speech Research Unit, DRA Malvern
St Andrews Road, Malvern, Worcs WR14 3PS, England

29th July 1992

Abstract

A simple statistical segmental approach to speech pattern modelling, based on segmental hidden Markov models, is proposed which addresses some of the limitations of conventional hidden Markov model based methods. The most important features of the new approach are the use of an underlying semi-Markov process to model speech at the segment level, rather than time-synchronous frame level, and to enable improved segment duration modelling, and the development of a segment model in which separate statistical processes are used to characterise extra-state and intra-state variability, thus making the temporal independence assumption more acceptable within a segment. A basic mathematical analysis of gaussian segmental hidden Markov models is presented and model parameter reestimation equations are derived. The relationship between the new type of model and variable frame rate analysis and conventional gaussian mixture based hidden Markov models is exposed.

INTENTIONALLY BLANK

Contents

1 Executive Summary 3

2 Introduction 4

3 Hidden Semi-Markov Models 6

3.1 Hidden Markov processes 6

3.2 Hidden semi-Markov processes 7

3.3 Advantages of Hidden Semi-Markov Models 8

4 Segmental Hidden Markov Models 8

4.1 Definition of the model 9

5 Gaussian Segmental HSMMs 10

5.1 A simple example 10

5.2 Mathematical analysis of Gaussian Segmental HMMs 12

5.2.1 Analysis of the State Model 12

5.2.2 Analysis of Multi-State Models 13

6 Relationship with Variable Frame Rate Analysis 14

6.1 Variable Rate Analysis 15

6.2 Improvements to the basic VFR algorithm 15

6.3 Interpretation of VFR analysis in terms of segmental HMMs 16

7 Relationship with multi-modal gaussian mixture densities 17

8 Parameter Re-estimation for Segmental HMMs 18

8.1 Derivation of the Reestimations Formulae 19

8.2 Remarks on the derivation of the reestimation formulae 24

9 Conclusions 24

A Proof of the concavity of $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ 27

B $\hat{\mathcal{Q}}(\mathcal{M},\bar{\mathcal{M}}) \rightarrow -\infty$ as $\bar{\mathcal{M}}$ approaches the boundary of the parameter space 29

B.1 Proof 29

B.2 Remarks 31

1 Executive Summary

Potential military applications of advanced speech technology are particularly demanding in terms of acoustic environment, channel characteristics, speaker variability and vocabulary flexibility. As higher-level techniques emerge which address these issues, ever-increasing demands are placed on recogniser performance at the acoustic-phonetic level. This is the fundamental stage in the recognition process where speech patterns derived from physical measurements are interpreted in terms of symbols which describe the basic sounds of the language. Performance at this level clearly depends on the quality of the speech models which are used.

The most successful automatic speech recognition systems use a statistical formalism, hidden Markov modelling, to model speech patterns, together with powerful mathematical methods for model parameter estimation and recognition. Hidden Markov models (HMMs) are currently the best compromise between mathematical tractability and acceptability from the perspective of speech science. However, from the latter perspective many of the assumptions which the HMM formalism makes about the structure of speech patterns are seriously in error. This constitutes a basic limitation on recogniser performance.

To overcome this limitation, new models are needed which more faithfully characterise important aspects of speech pattern structure and which, at the same time, are amenable to rigorous mathematical methods for parameter estimation and classification.

From the viewpoint of speech pattern modelling, the most significant limitations of the conventional HMM formalism are:

(i) the time-synchronous nature of the modelling, where it is assumed that the acoustic feature vector at a particular time depends only on the state of the model at that time and is otherwise independent of the preceeding vectors

(ii) the assumption that speech patterns are piecewise stationary with instantaneous transitions bewteen stationary regions.

This report describes an extension of the HMM formalism which tackles the first limitation by taking explicit account of the segmental nature of speech patterns. This is seen as a step towards the development of dynamic segmental statistical models which are able to explicitely model the dynamic behaviour of speech patterns.

An initial theory of time-asynchronous static segmental HMMs is presented, in which sources of extra-segmental variation, such as identity of speaker or the particular choice of acoustic target, are fixed throughout a segment rather than being allowed to vary time-synchronously as in a conventional HMM. The most important result is that the conventional HMM parameter estimation algorithm can be extended to this new type of segmental HMM. The main part of the memorandum is concerned with the derivation of the extended parameter estimation algorithm and a formal proof of its validity.

INTENTIONALLY BLANK

2 Introduction

At present the most successful automatic speech recognition systems, in terms of recognition accuracy, are those which use hidden Markov models (HMMs) to model speech at the acoustic level and dynamic programming based recognition algorithms which find the best interpretation of an unknown speech pattern in terms of the output of a sequence of HMMs. The most recent systems, such as those developed under the DARPA Spoken Language Systems project and at IBM and Dragon in the USA, and RSRE's "ARMADA" system in the UK, use HMMs to model speech at the phoneme level in order to address medium to large vocabularies and to avoid vocabulary-specific training.

This success is due to two factors. Firstly HMMs provide a formal statistical framework which is broadly appropriate for modelling speech patterns. This single framework is able simultaneously to accomodate the time-varying nature of speech patterns, through the structure of the underlying Markov process, and the variable segmental structure of these patterns through the statistical processes which are identified with the states of the model. Secondly there exist computationally useful and rigorous mathematical methods for automatically optimising the parameters of a set of HMMs relative to training data, and for classifying an unknown speech pattern given a set of HMMs. These are the Baum-Welch algorithm, which is used to adjust the parameters of a set of HMMs in order to (locally) maximise the probability of a given set of training material conditioned on these HMMs, and the Viterbi, or One-Pass Dynamic Programming algorithm which computes, in a particular sense, the most probable sequence of HMMs given an unknown speech pattern.

These two factors taken together (a broadly appropriate formalism and the existence of rigorous mathematical methods for manipulating that formalism) constitute a powerful tool for speech pattern processing. However from the perspective of speech science it is clear that the assumptions which the HMM formalism imposes on the structure of speech patterns are inappropriate in several respects.

(a) **Piecewise Stationarity** The HMM framework assumes that a speech pattern is produced by a piecewise stationary process, with instantaneous transitions between the stationary states. This is clearly at variance with the fact that speech patterns are derived from signals produced by a continuously moving physical system - the vocal tract.

(b) **Properties of the States** In a standard HMM the statistical process associated with a state is defined by a single probability density function (pdf). This pdf typically has to accommodate several quite distinct types of variability, for example:

- Long-term extra-segmental variations, such as speaker sex, identity of speaker, and long-term prosodic phenomena, which are essentially fixed throughout the duration of a segment.
- Short-term intra-segment variations which occur once the segment target has been achieved.

In addition, in reality the configuration of the vocal tract is not even nominally stationary, for example in the dynamic part of a diphthong or in most consonants, and this is another source of variablity.

A further consideration is the fact that in order for Baum-Welch parameter reestimation theory to apply, the class of the state output pdf is resricted to non-parametric discrete distributions (in the case where the front-end processing includes quantisation to ensure that all observation vectors are drawn from given a finite set), or mixtures of multivariate gaussian pdfs (see [7]). The extent to which such pdfs are appropriate for modelling acoustic feature vectors in speech patterns has been considered by Richter [9]

(c) **The Independence Assumption** It is assumed that the probability that a given acoustic vector corresponds to a given state of the HMM depends (directly) only on the vector and the state, and is otherwise independent of the sequence of acoustic vectors and states which preceed and succeed the current vector and state. Thus the model take no account of the dynamical constraints of the physical system which has generated a particular sequence of acoustic data.

Clearly, the problems associated with the independence assumption are exacerbated by the use of a single density to model all sources of variability (see (b)). For example, in a speaker-independent system in which high-order mixture densities are used to model inter-speaker variations, the model assumes that each acoustic feature vector in a sequence may have been produced by a different speaker.

(d) **State Duration** Because of the Markov assumption, state (and hence speech segment) duration in a HMM conforms to a geometric pdf which assigns maximum probability to state duration 1 and successively smaller probabilities to longer durations. This is not an appropriate model of speech segement duration.

(e) **Model Topologies** The basic segmental-sequential structure of the patterns corresponding to a particular HMM is determined by the topology of the underlying Markov model (i.e. the number of states and the permitted transitions between states). In most HMM-based speech recognition systems a common HMM topology is chosen for all models. However the patterns which are to be modelled typically exhibit a range of types of sequential strucure.

Most of the progress which has been achieved over recent years has resulted from working within this basic HMM framework. There has been very little work aimed at extending the HMM formalism in ways which address the limitations listed above. One reason for this is the realisation of the importance of the mathematical tools associated with HMMs and an acknowledgement of the need to extend these mathematical techniques in parallel with any extension of the basic formalism.

An example of the way in which the conventional HMM formalism can be extended is the work on hidden *semi*-Markov markov models (HSMMs) reported in [10], [11] and [6] in which the HSMM structure enables the geometric model of state duration in a standard HMM to be replaced by something more appropriate. In a HSMM the underlying Markov process is replaced with a semi-Markov process in which state duration is explicitely modelled by state-dependent state duration pdfs. It was shown that the standard HMM optimisation and recognition algorithms can be extended to HSMMs with non-parametric, Poisson and Gamma state duration pdfs [6, 10]. Small vocabulary speech recognition experiments were conducted which showed that HSMMs could consistently outperform standard HMMs [3, 12].

The essential difference between HMMs and HSMMs is that HMMs are time-synchronous in the sense that states are associated with single acoustic vectors, whereas in a HSMM states are associated with sequences of acoustic vectors. Hence, in addition to their utility for duration modelling, HSMMs offer a computationally useful framework for more general modelling of speech at the segment level.

The purpose of this memorandum is to present a new HSMM based segment level stochastic model which addresses some of the limitations of HMMs which have been listed above, and which at the same time is computationally useful in the sense that the existing HMM parameter estimation and classification algorithms can be extended to this new class of model. The basis of the new model is the notion of separating the modelling of sources of variability which apply above the segment level from that of sources of variability which apply within a segment. Intuitively, since in the present context segments are sub-phonemic, variations due to factors ranging from identity of speaker down to the choice of "target realisation" of a particular sound fall into the first category, while subsequent variations around that target fall into the second category. It is this perspective which leads to the use of terms such as "target" in the discussions which follow.

The organisation of the memorandum is as follows. Section 3 introduces the terminology and notation of hidden Markov and semi-Markov models which is necessary for the development of segmental hidden semi-Markov models. Section 4 presents a general formal definition of this new type of segmental model. Section 5 introduces the special case of gaussian segmental HSMMs. A simple example of this type of model is compared with the corresponding conventional HSMM. The section goes on to present a basic mathematical analysis of gaussian segmental HSMMs. In section 6 it is shown that gaussian segmental models can be viewed as an extension of conventional variable frame-rate analysis in which dynamic programming based variable frame rate analysis is integrated with Markov model based processing. The relationship with HMMs with gaussian mixture densities is explored in section 7. Section 8 presents a derivation of Baum-Welch type reestimation formulae for segmental HSMM parameters.

3 Hidden Semi-Markov Models

3.1 Hidden Markov processes

In the standard hidden Markov model (HMM) based approach to speech pattern modelling it is assumed that a sequence of observed multi-dimensional acoustic vectors,

$$y = y_1, y_2, \ldots, y_t, \ldots, y_T$$

corresponding to a given speech signal, is a probabilistic function of a hidden state sequence

$x = x_1, x_2, ..., x_t, ..., x_T$

where each x_t is drawn from a finite set of states $\sigma = \{\sigma_1, .., \sigma_N\}$. The sequential and durational statistics of x are determined by a transition probability matrix

$A = [a_{ij}]_{i,j=1,...,N}$

where, $a_{ij} = Prob(x_t = \sigma_j | x_{t-1} = \sigma_i)$ is the probability of a transition from state σ_i to state σ_j, and an initial state probability vector

$\pi = [\pi_i]_{i=1,...,N}$

where $\pi_i = Prob(x_1 = \sigma_i)$. The pair $\mathcal{M} = (\pi, A)$ define an N state Markov process. The relationship between the observation vectors y_t and the hidden states x_t is defined by a set of probability density functions $\{b_i\}_{i=1,...,N}$, where

$b_i(o) = Prob(y_t = o | x_t = \sigma_i)$

is the probability that the observation o is associated with state σ_i. The triple $\mathcal{H} = (\pi, A, \{b_i\})$ defines a hidden Markov process. The process is called hidden because it is not possible to infer unambiguously the exact state sequence which gave rise to a particular observation sequence.

3.2 Hidden semi-Markov processes

A semi-Markov process is obtained by associating a probability density function $\mathcal{D}_i$, defined on the set of positive integers, with each state σ_i of an N-state Markov process $\mathcal{M} = (\pi, A)$. For $d = 1, 2, 3, .., \mathcal{D}_i(d)$ is the probability of occupying state σ_i for precisely d time units. The density $\mathcal{D}_i$ is called the state duration pdf associated with state σ_i, and the Markov process $\mathcal{M}$ is called the underlying Markov process.

A hidden semi-Markov process is a probabilistic function of a semi-Markov process. More precisely, an N state hidden semi-Markov model (HSMM), or Variable Duration HMM [6], is a 4-tuple $\mathcal{S} = (\pi, A, \{\mathcal{D}_i\}, \{b_i\})$ where:

- $\mathcal{M} = (\pi, A)$ is an N-state Markov model
- $\mathcal{D}_1, ..., \mathcal{D}_N$ is a set of N state duration pdfs, $\mathcal{D}_i : \mathbf{N} \rightarrow [0, 1]$
- $b_1, ..., b_N$ is a set of N state output pdfs, $b_i : \mathbf{R}^d \rightarrow [0, 1]$

where $\mathbf{N}$ and $\mathbf{R}^d$ denote the positive integers and real d dimensional space respectively.

Intuitively one can visualise a hidden semi-Markov process as follows. At some time $t = 1$ the process enters state $x_1 = \sigma_i$, chosen randomly according to the initial state probability vector π. A duration d_1 is chosen randomly according to the state duration pdf $\mathcal{D}_i$, and a sequence $y_1, ..., y_{d_1}$ of d_1 acoustic vectors is generated randomly and independently according to the state output pdf b_i. The process then moves from state σ_i

to state σ_j according to the state transition probability matrix A. In general, at time $t = d_1 + d_2 + ... + d_{m-1} + 1$ the process enters state $x_m = \sigma_i$. As in the case of the first state, a duration d_m is chosen randomly accordiug to the state duration pdf $\mathcal{D}_i$, and a sequence $y_t, ..., y_{t+D_m-1}$ of d_m acoustic vectors is generated randomly and independently according to the state output pdf b_i.

It is straightforward to show the principle of dynamic programming can be extended from Markov to semi-Markov processes. Consequently the standard dynamic programming based recognition algorithms can be extended from HMMs to HSMMs. In addition it has been demonstrated that Baum's theorem (and hence the Baum-Welch parameter estimation algorithm) can be extended to HSMMs with discrete, Poisson or Gamma state duration pdfs ([10, 6]). In all cases the need to explicitely consider times $t - \delta$ ($\delta = 1, 2, ..., d_{max}$) during HSMM based computations leads to an increase in computational load relative to HMMs, however HSMMs still provide a computationally useful formalism.

3.3 Advantages of Hidden Semi-Markov Models

In the past, HSMMs have primarily been used to remedy the limitations of HMMs with respect to speech segment duration modelling, and have not been used to address any of the other limitations of HMMs which are listed in the introduction. Consequently, because the improvements in recognition accuracy which result from better duration modelling are generally relatively modest and the increase in computational load is relatively high, there has been little recent work in this area. The objective of this memorandum is to show that, leaving the duration modelling capabilities of HSMMs aside, the segment based formalism provided by HSMMs can be exploited to address some of the other limitations of HMMs.

4 Segmental Hidden Markov Models

This section proposes a segmental model of speech which is an extension of the conventional HSMM described above. The model is motivated by the need to explicitely deal separately with the different types of variability which are accomodated in the state output pdf of a conventional HMM or HSMM, thereby making the independence assumption more realistic. Hence the new model explicitely addresses points (b) and (d) of the introduction and implicitely addresses point (c).

In a conventional HSMM the stochastic process associated with state σ_i is defined by a state output pdf

$b_i : \mathbf{R}^n \to [0, 1]$.

In the associated model of speech pattern production, at each time t an observation vector y_t is produced randomly according to the pdf b_i. The vector y_t clearly depends on the state σ_i but is otherwise independent of the observations $y_1, ..., y_{t-1}$ which preceded it. However, it has already been noted that the pdf b_i typically accomodates several different types of variability, including variations in the target for a given sound (both intra-speaker and inter-speaker) and natural variations which occur once the target has been chosen. Because all of these types of variablity are modelled by a single pdf and the sequence of observation vectors are generated independently (according to that pdf) there is nothing to prevent successive observation vectors corresponding to quite different sets of extra-segmental factors, such as different speakers.

The proposed model overcomes this problem by using separate processes to model extra-state and intra-state variations.

4.1 Definition of the model

In the state model described below, extra-state variations associated with state σ_i are modelled by a pdf b_i called the *state target pdf*. Arrival at state σ_i causes a single output to be generated by the process associated with this pdf. This output, which will be called a *target* is a pdf v which can be regarded as modelling within-state variability once all sources of extra-state variability have been fixed. Thus on entering state σ_i, a state duration d is chosen randomly according to the state duration pdf $\mathcal{D}_i$ and a target v is chosen randomly according to the pdf b_i. A sequence of d observation vectors is then produced, with each individual observation being generated randomly and independently of its successors according to the pdf v.

More formally, the stochastic process associated with state σ_i is governed by a probability density function

$b_i : \mathcal{P}^n \rightarrow [0,1]$

where $\mathcal{P}^n$ denotes a subset of the set of probability density functions defined on n-dimensional space $\mathbf{R}^n$.

In other words, in the new type of model a *target* is defined to be a pdf v which can be thought of as modelling variations in the acoustic pattern which occur once the speaker and all other sources of extra-state variability have been selected. This target is fixed throughout a particular state occupancy. The *state target pdf* b_i specifies the probability of any particular target given state σ_i.

Hence although the sequence of observation vectors generated by such a state model are still independent samples from the same distribution v, in this case the distribution v only models variations which occur given a fixed acoustic target. All of the vectors which are generated during a particular state occupancy are constrained to correspond to the same target.

Therefore, formally, an N-state segmental HSMM is a 5-tuple $\mathcal{M} = (\pi, A, \{\mathcal{D}_i\}, \mathcal{P}^n, \{b_i\})$ where

- $(\pi, A, \{\mathcal{D}_i\})$ $(i = 1, .., N)$ is an N-state semi-Markov model
- $\mathcal{P}^n$ is a set of target pdfs defined on n-dinensional space R^n

- $b_i : \mathcal{P}^n \rightarrow [0,1]$ is a state target pdf defined on $\mathcal{P}^n$

In the above notation, a state of a segmental HSMM is a triple

$\sigma = (\mathcal{P}^n, b, \mathcal{D})$

Given a sequence of observations

$y = y_1, ..., y_t, ..., y_T$

the joint probability of the sequence y and a particular target pdf $v : \mathbf{R}^n \rightarrow [0,1]$ given σ is therefore given by:

$$P_\sigma(y, v) = \mathcal{D}(T) b(v) \prod_{t=1}^{T} v(y_t) \tag{1}$$

5 Gaussian Segmental HSMMs

Suppose that a target is defined to be any gaussian pdf defined on n-dimensional space $\mathbf{R}^n$ with fixed variance τ. Given that the variance is fixed, a target is defined uniquely by its mean and hence in this case $\mathcal{P}^n$ is equal to $\mathbf{R}^n$. For a given state σ_i let the state target pdf b_i be a gaussian pdf $\mathcal{N}_{(\mu_i, \gamma_i)}$ defined on $\mathcal{P}^n = \mathbf{R}^d$ with mean μ_i and variance γ_i.

5.1 A simple example

To illustrate this, consider the simple case of a 1-state 1-dimensional model with state mean $\mu_1 = 0$ and variance $\gamma_1 = 6$, and with fixed target variance $\tau = 0.5$. In addition, suppose that state duration follows a Poisson distribution with mean state duration $\delta_1 = 300$,. Figure 1 shows a sequence of 1100 random observations generated by such a process in the manner described in section 4.1. The figure clearly shows 3 separate state occupancies. At the start of each occupancy a target value and a duration are chosen and the appropriate number of observations are generated, at random, from a relatively tight pdf centered about the target.

Now consider the result of trying to model this as a conventional 1-state hidden semi-Markov process. The latter assumes that all of the observations are generated randomly and independently by a single gaussian process. It is straightforward to show that the mean of this process is equal to 0, the mean of the state target pdf of the segmental process, and its variance is 0.6, the sum of the variances of the state target pdf and individual target pdfs of the segmental process. Figure 2 shows a sequence of 1100 observations from such a process given the correct Poisson duration statistics. State transitions occur at the same times as in figure 1, however all of the structure apparent in figure 1 which shows the separate state occupancies has been lost.

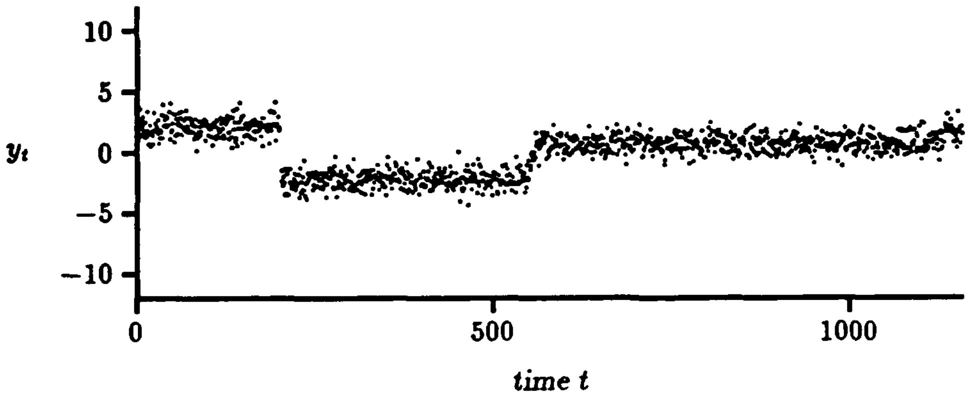

Figure 1: *Observations from 1 state segmental hidden semi-Markov process with Poisson state duration statistics*

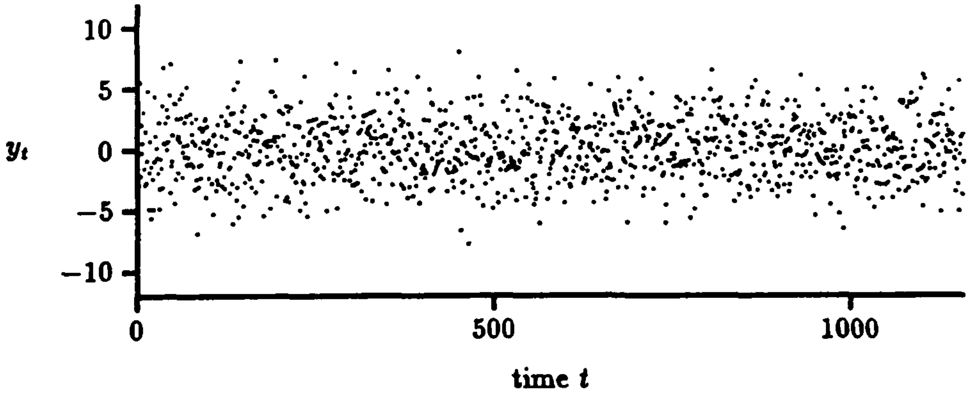

Figure 2: *Observations from 1 state hidden semi-Markov process with Poisson state duration statistics*

5.2 Mathematical analysis of Gaussian Segmental HMMs

The purpose of this section is to present the basic equations which are necessary for the study of gaussian segmental HSMMs. In order to focus attention onto those aspects which are relevant to the segmental nature of the models, three simplifying assumptions have been made. First, the precise form of the state duration pdf associated with a particular state σ_i is not specified. This pdf is simply denoted by $\mathcal{D}_i$, and it is assumed that it is independent of the parameters of the segment model, namely μ_i, γ_i and τ_i $(i = 1, ..., N)$. Second it is assumed that all observations are 1-dimensional. The latter assumption is also unnecessary, and it will be seen that the arguments can be extended to multi- dimensional observations, but is made for the reasons given above. The final simplifying assumption is that the underlying Markov model is strictly left-right. Again this is not necessary but it will significantly simplify notation, particularly in the derivation of the parameter reestimation formulae in section 8.

5.2.1 Analysis of the State Model

In the above notation, a state σ of a gaussian segmental HSMM is a triple

$\sigma = (\mathcal{P}, \mathcal{N}_{(\mu,\gamma)}, \mathcal{D})$

where $\mathcal{P}$ is the set of pdfs defined on $\mathbf{R}$ of the form $\mathcal{N}_{(x,\tau)}$ $(x \in \mathbf{R})$, and $\mathcal{D}$ denotes an appropriate state duration pdf. Given a sequence of observations

$y = y_1, ..., y_t, ..., y_T,$

the joint probability of the observation sequence y and a particular target[1] c given state σ is given by

$$P_\sigma(y, c) = \mathcal{D}(T)\mathcal{N}_{(\mu,\gamma)}(c) \prod_{t=1}^{T} \mathcal{N}_{(c,\tau)}(y_t) \tag{2}$$

and the probabilty of the sequence y given σ is

$$P_\sigma(y) = \int_c P_\sigma(y, c) \tag{3}$$

An alternative to the "full probability" criterion of equation (3), which is more ammenable to analysis, is to consider the joint probability $P_\sigma(y, \hat{c})$, where $\hat{c}$ is the value of the target c which maximises $P_\sigma(y, c)$. Define

$$\begin{aligned} \hat{P}_\sigma(y) &= max_c P_\sigma(y, c) & (4) \\ \hat{c} &= argmax_c P_\sigma(y, c) & (5) \end{aligned}$$

Claim

[1] Here the term "target" is being used to refer either to the gaussian pdf $\mathcal{N}_{(c,\tau)}$ with mean c and fixed variance τ or the mean value c

Let μ, γ, τ and y be as above, then

$$\hat{c} = \frac{\mu\tau + \sum_{t=1}^{T} y_t\gamma}{\tau + T\gamma} \tag{6}$$

Proof

Since the logarithm function is monotonic it is enough to show that $\hat{c}$ maximises $log(P_\sigma(y,c))$. From equation (2),

$$\begin{aligned} log(P_\sigma(y,c)) &= log(\mathcal{D}(T)) + log(\mathcal{N}_{(\mu,\gamma)}(c)) \\ &\quad + \sum_{t=1}^{T} log(\mathcal{N}_{(c,\tau)}(y_t)) \end{aligned} \tag{7}$$

Therefore,

$$\begin{aligned} \frac{\partial}{\partial c} log(P_\sigma(y,c)) &= \frac{\partial}{\partial c} log\mathcal{D}(T) \\ &\quad + \frac{\partial}{\partial c} log(\mathcal{N}_{(\mu,\gamma)}(c)) \\ &\quad + \sum_{t=1}^{T} \frac{\partial}{\partial c} log(\mathcal{N}_{(c,\tau)}(y_t)) \\ &= \frac{(\mu - c)}{2\gamma} - \sum_{t=1}^{T} \frac{(c - y_t)}{\tau} \\ &= \frac{\mu\tau + \sum_{t=1}^{T} y_t\gamma - c(\tau + T\gamma)}{\gamma\tau} \end{aligned} \tag{8}$$

Setting the right-hand side of equation (8) to zero, multiplying through by $\gamma\tau$, and solving for c gives the required result. To see that (6) defines a maximum, note that from (8):

$$\frac{\partial^2}{\partial c^2} log(P_\sigma(y,c)) = -\frac{(\tau + T\gamma)}{\gamma\tau} < 0 \tag{9}$$

because γ and τ are both positive.

Equation (6) has interesting properties. The expression for $\hat{c}$ is a linear combination of μ, the expected output of state σ and the sum $\sum_{t=1}^{T} y_t$ of the observations. If the variance τ of the target is large, so that the observations are not expected to be tightly constrained by the target process, then $\hat{c}$ is biased towards the state mean μ. However, if the state variance γ is large and τ is small then $\hat{c}$ is biased towards the actual observations.

5.2.2 Analysis of Multi-State Models

Now consider an N state segmental HMM $\mathcal{M}$, where the ith state σ_i of $\mathcal{M}$ is defined by

$\sigma_i = (\mathcal{N}_{(\mu_i,\gamma_i)}, \mathcal{N}_{(.,\tau_i)}, \mathcal{D}_i)$

To simplify mathematical notation in what follows it will be assumed that the underlying Markov process for $\mathcal{M}$ is strictly left-right, in the sense that if $t \geq s$, $x_t = \sigma_j$ and $x_s = \sigma_i$

then $j \geq i$. This assumption is not necessary, but it enables a state sequence x to be written in the form

$x = d_1 \otimes \sigma_1, ..., d_i \otimes \sigma_i, ..., d_N \otimes \sigma_N$

where $d_i \otimes \sigma_i$ denotes duration d_i in state σ_i. Without this assumption it is necessary to introduce an extra level of indirection, to map the mth state visited in the sequence x onto the correct σ_i and to account for multiple occurances of states in the sequence x. This is straightforward, but the additional notation which is required obscures the basic simplicity of the ideas which foillow.

The joint probability $P(y, x|\mathcal{M})$ of the observation sequence y and the state sequence x given the model $\mathcal{M}$ is given by

$$P(y, x|\mathcal{M}) = \prod_{i=1}^{N} a_{i-1,i} P_{\sigma_i}(y_{t_{i-1}+1}^{t_i}) \tag{10}$$

where $y_s^t = y_s, y_{s+1}, ..., y_t$, and t_i is the largest value of t for which $x_t = \sigma_i$ $(i \geq 1)$, $t_0 = 0$.

The probability of y given the model $\mathcal{M}$ is then given by:

$$P(y|\mathcal{M}) = \sum_x P(y, x|\mathcal{M}) \tag{11}$$

By analogy, define

$$\begin{aligned} \hat{P}(y, x|\mathcal{M}) &= \prod_{i=1}^{N} a_{i-1,i} \hat{P}_{\sigma_i}(y_{t_{i-1}+1}^{t_i}) && (12) \\ \hat{P}(y|\mathcal{M}) &= \sum_x \hat{P}(y, x|\mathcal{M}) && (13) \end{aligned}$$

Hence $\hat{P}(y, x|\mathcal{M})$ is similar to the joint probability $P(y, x|\mathcal{M})$ except that in the computation of $\hat{P}(y, x|\mathcal{M})$ the evaluation of the probability of a particular subsequence of y given a state σ_i is based on the optimal target $\hat{c}$. Note that since $\hat{c}$ depends on the state sequence x and the state σ_i it is more correct to write

$$\hat{c}_{x,i} = \frac{\mu_i \tau_i + \sum_{t=t_{i-1}+1}^{t_i} y_t \gamma_i}{\tau_i + d_i \gamma_i} \tag{14}$$

The analysis which follows, and in particular the derivation of reestimation formulae in section 8, will focus on the quantities $\hat{P}(y, x|\mathcal{M})$ and $\hat{P}(y|\mathcal{M})$ rather than $P(y, x|\mathcal{M})$ and $P(y|\mathcal{M})$

6 Relationship with Variable Frame Rate Analysis

In this section it will be shown that the segmental HMM based analysis proposed in this memorandum can be regarded as an extension and intergration of conventional Variable Frame Rate (VFR) analysis and hidden Markov modelling

6.1 Variable Rate Analysis

Variable frame rate (VFR) analysis is a method for data-rate reduction which has been shown to give improved performance over fixed frame rate analysis for automatic speech recognition [8]. In its simplest form VFR is used to remove vectors from an observation sequence. A distance is computed between the current observation vector and the most recently retained vector, and the current vector is discarded if this distance falls below a threshold T. When a new observation vector causes the distance to exceed the threshold, the new vector is kept and becomes the most recently retained vector. VFR analysis replaces sequences of similar vectors with a single vector, and hence reduces the amount of computation required for recognition.

What is interesting is that VFR analysis can also improve recognition accuracy [8]. There are a number of possible explanations for this:

- by discarding vectors from relatively stationary regions of the speech pattern, VFR focusses the recognition process onto the dynamic regions, which are important for classification
- vectors in the relatively stationary regions of speech patterns are highly correlated, contrary to the assumption of independence which is part of the HMM formalism. Discarding vectors in these regions results in observation sequences which are more consistent with the formalism
- if a count of the number of frames which each retained vector replaces is appended to that vector, then some implicit duration modelling is incorporated into the recognition process

6.2 Improvements to the basic VFR algorithm

The basic VFR algorithm described above can be improved in a number of ways:

6.2.1 Rather than replacing a sequence of acoustic vectors $y_s, ..., y_t$ with y_s, the first vector in the sequence, it should replaced with some form of average $\bar{y}_s^t$ taken over the sequence.

6.2.2 For a finite sequence $y = y_1, ..., y_T$ the "left-right" threshold based segmentation method used in the basic VFR algorithm should be replaced with a "global" dynamic programming based segmentation algorithm such as that described in [4]. The dynamic programming based method does not rely on a threshold. It is used to partition the sequence y into a sequence of M subsequences $y_1^{t_1}, ..., y_{t_{i-1}+1}^{t_i}, ..., y_{t_{M-1}+1}^{t_M}$ $(1 \leq t_1 \leq ... \leq t_M = T)$ such that some criterion

$$Dist(t_1, ..., t_i, ..., t_M) = \sum_{i=1}^{M} D(y_{t_{i-1}+1}^{t_i}) \tag{15}$$

is minimised. The quantity $D(y^{t_i}_{t_{i-1}+1})$ is typically a distortion measure on the sequence $y^{t_i}_{t_{i-1}+1}$, for example the sum of euclidean distances between vectors in the sequence and the sequence mean.

6.2.3 In the context of Markov model based speech pattern processing it is clearly sub-optimal to segment the sequence of acoustic observation vectors and discard information during VFR analysis, and then to perform a second state-level segmentation. The segmentation of the observation sequence during VFR analysis should be integrated with the state-level segmentation performed in the model based analysis.

6.3 Interpretation of VFR analysis in terms of segmental HMMs

It will be shown that extending the basic VFR analysis algorithm in the ways described above leads naturally to a segmental HMM based analysis. Suppose that $\mathcal{M} = (\pi, A, \{b_i\})$ is a HMM, with $b_i = \mathcal{N}_{(\mu_i,\gamma_i)}$, and that $y = y_1, ..., y_t, ..., y_T$ is a sequence of acoustic vectors in R^d. In a dynamic programming based VFR scheme of the type alluded to in 6.2.2 above, dynamic programming is used to find a partition of the sequence y into M subsequences $y^{t_1}_1, ..., y^{t_i}_{t_{i-1}+1}, ..., y^{t_M}_{t_{M-1}+1}$, such that

$$Dist(t_1, ..., t_i, ..., t_M) = \sum_{i=1}^{M} D(y^{t_i}_{t_{i-1}+1}) \tag{16}$$

is minimised.

Taking account of (6.2.1), following VFR analysis the sequence y would be represented by the sequence

$\bar{y} = \bar{y}^{t_1}_1, ..., \bar{y}^{t_i}_{t_{i-1}+1}, ..., \bar{y}^{t_M}_{t_{M-1}+1}$

where $\bar{y}^{y_i}_{t_{i-1}+1}$ denotes some form of average over the sequence $y^{y_i}_{t_{i-1}+1}$.

During subsequent HMM based processing, dynamic programming is used again to find a state sequence $x = x_1, ..., x_M$ relative to the HMM $\mathcal{M}$, such that the probability

$$P(\bar{y}, x|\mathcal{H}) = \prod_{i=1}^{M} a_{x_{i-1},x_i} \mathcal{D}_{x_i}(d_i) b_{x_i}(\bar{y}^{t_i}_{t_{i-1}+1}) \tag{17}$$

is maximised. Here $\mathcal{D}_{x_i}$ is a state dependent duration pdf which is applied to the VFR count d_i.

The goal of 6.2.3 is that ideally the two equations (16) and (17) should be optimised simultaneously rather than separately. To achieve this it is necessary to make assumptions about the form of the distortion measure D. Suppose that

$$D(y^{t_i}_{t_{i-1}+1}) = \sum_{t=t_{i-1}+1}^{t_i} D_{EUC}(y_t, \bar{y}^{t_i}_{t_{i-1}+1}) \tag{18}$$

where D_{EUC} denotes the squared euclidean metric. Then, since

$$D_{EUC}(y_t, \bar{y}^{t_i}_{t_{i-1}+1}) = -K_1 log(\mathcal{N}_{(\bar{y}^{t_i}_{t_{i-1}+1}, 1)}(y_t) + K_2 \tag{19}$$

where K_1 and K_2 are constants, minimising equation (16) is equivalent to maximising the quantity

$$P(t_1, ..., t_i, ..., t_M) = \prod_{i=1}^{M} \prod_{t=t_{i-1}+1}^{t_i} \mathcal{N}_{(\bar{y}^{t_i}_{t_{i-1}+1}, 1)}(y_t) \tag{20}$$

Equations (17) and (20) can now be combined to give an evaluation criterion for a VFR analysis scheme which satisfies 6.2.1, 6.2.2 and 6.2.3:

$$P(\bar{y}, x|\mathcal{H}) = \prod_{i=1}^{M} a_{x_{i-1},x_i} \mathcal{D}_{x_i}(d_i) b_{x_i}(\bar{y}^{t_i}_{t_{i-1}+1}) \prod_{t=t_{i-1}+1}^{t_i} \mathcal{N}_{(\bar{y}^{t_i}_{t_{i-1}+1}, 1)}(y_t) \tag{21}$$

But equation (21) has precisely the same form as equation (12), with $\tau_i = 1$, for all i, and

$$\bar{y}^{t_i}_{t_{i-1}+1} = \hat{c}_{x,i} \tag{22}$$

In other words, replacing the basic VFR analysis procedure described in section 6.1 with the obvious dynamic programming based method and then integrating this with the higher-level Markov model based processing leads naturally to the type of gaussian segmental HMM based analysis which is proposed in this memorandum. In this sense segmental HMMs can be regarded as a natural extension and integration of VFR analysis and HMM-based analysis.

7 Relationship with multi-modal gaussian mixture densities

One of the classes of state output pdf which is frequently used in conventional hidden Markov modelling is the class of gaussian mixture densities. In such a system the state output pdf b_i associated with the ith state has the form

$$b_i(o) = \sum_{j=1}^{J} w_j \mathcal{N}_{(\mu_j, \gamma_j)}(o) \tag{23}$$

for any observation o, where $\sum_{j=1}^{J} w_j = 1$. There is also a continuous analogue of (23) of the form

$$b_i(o) = \int_j w(j) \mathcal{N}_{(\mu_j, \gamma_j)}(o) dj \tag{24}$$

where $\int_j w(j)dj = 1$ Parameter reestimation formulae for models based on (23) and (24) have been established in [7] and [5], and in [7] respectively.

Gaussian mixture state models are used to compensate for the fact that in reality the set of observations associated with a particular state will not generally be consistent with a single gaussian process. This is particularly true in cases where the models in question are used to characterise speech from a number of speakers. Thus, gaussian mixtures are typically used to model broad sources of extra-segmental variablity and hence, from the viewpoint of this memorandum, they exacerbate the problems associated with the independence assumption within a state.

The segment model proposed here is clearly related to (24), however in the new type of model a single component of the continuous mixture is chosen on entering a state and all observations emitted during a particular state occupancy are drawn from that component. In the case of (24) a different component can be used to explain each individual observation. Thus, the new type of model can be regarded as a continuous gaussian mixture in which all observations corresponding to a particular state occupancy are constrained to come from the process associated with a single component of that mixture.

8 Parameter Re-estimation for Segmental HMMs

The analysis presented in this section is concerned with the derivation of reestimation formulae for the parameters of the state segment models, namely μ_i, γ_i and τ_i. The reestimation formulae for the initial state probabilities π_i and state transition probabilities $a_{i,j}$ are the same as those presented in [10] and are not re-derived here. Similarly, no precise form for the state duration pdfs is assumed other than that they should be independent of the parameters of the state segment models. Reestimation formulae for non-parameteric discrete, poisson and gamma state duration pdfs are given in [10, 6].

As in section 5.2, the derivations presented below are simplified by assuming that all observation vectors y_t are 1-dimensional and that the underlying Markov model is strictly left-right. Again it is emphasised that these assumptions are not necessary but are made in order to focus attention onto those aspects of the mathematics which are directly relevant to the segmental nature of the models and to reduce notational complexity.

The analysis focuses on the quantity $\hat{P}(y|\mathcal{M})$. Hence, given a model $\mathcal{M}$, the goal of reestimation is to derive a new model $\bar{\mathcal{M}}$ such that

$$\hat{P}(y|\bar{\mathcal{M}}) \geq \hat{P}(y|\mathcal{M}) \tag{25}$$

As for conventional HMMs and HSMMs [2, 7] it is convenient at this point to introduce a function $\hat{Q}(\mathcal{M},.)$ of $\bar{\mathcal{M}}$, called the auxiliary function, defined by,

$$\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) = \sum_x \hat{P}(y,x|\mathcal{M}) log \hat{P}(y,x|\bar{\mathcal{M}}) \tag{26}$$

The auxiliary function has the following property

Claim

If $\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) \geq \hat{Q}(\mathcal{M},\mathcal{M})$ then $\hat{P}(y|\bar{\mathcal{M}}) \geq \hat{P}(y|\mathcal{M})$

Proof

The proof is the same as in [2, 7]

It follows that in order to find a model $\bar{\mathcal{M}}$ which satisfies equation (25) it is sufficient to find $\bar{\mathcal{M}}$ such that $\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) > \hat{Q}(\mathcal{M},\mathcal{M})$. In particular it is sufficient to find a model $\bar{\mathcal{M}}$ which maximises $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$. This maximum is obtained by setting the partial derivatives of $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ with respect to the parameters of $\bar{\mathcal{M}}$ equal to zero and solving the resulting equations.

8.1 Derivation of the Reestimations Formulae

Claim

Let y be a sequence of observation vectors and let $\mathcal{M}$ be a gaussian segmental HMM as in section 5.2.2. Let $\bar{\mathcal{M}}$ be the gaussian segmental HMM with parameters defined as follows,

$$\bar{\mu}_i = \frac{\sum_{x\in S_i} P(y,x|\mathcal{M}) \sum_{t=t_{i-1}+1}^{t_i} y_t}{\sum_{x\in S_i} P(y,x|\mathcal{M}) d_i} \tag{27}$$

$$\bar{\gamma}_i = \frac{\sum_{x\in S_i} P(y,x|\mathcal{M}) (\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2}{\sum_{x\in S_i} P(y,x|\mathcal{M})} \tag{28}$$

$$\bar{\tau}_i = \frac{\sum_{x\in S_i} P(y,x|\mathcal{M}) \sum_{t=t_{i-1}+1}^{t_i} (\hat{\bar{c}}_{x,i} - y_t)^2}{\sum_{x\in S_i} P(y,x|\mathcal{M}) d_i} \tag{29}$$

where $S_i = \{x : x_t = \sigma_i \text{ for some } t\}$ and

$$\hat{\bar{c}}_{x,i} = \frac{\bar{\mu}_i \bar{\tau}_i + \sum_{t=t_{i-1+1}}^{t_i} y_t \bar{\gamma}_i}{\bar{\tau}_i + d_i \bar{\gamma}_i} \tag{30}$$

Then provided that

(i) $\bar{\gamma}_i > \bar{\tau}_i$ for all i, and

(ii) the sequence $y = y_1, ..., y_T$ is not constant

$\hat{P}(y|\bar{\mathcal{M}}) \geq \hat{P}(y|\mathcal{M})$.

Proof

The arguments follow those in [2] and [7].

From the previous discussion it is sufficient to show that the model $\bar{\mathcal{M}}$ defined above maximises $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ as a function of $\bar{\mathcal{M}}$. The proof is divided into three stages:

- $\bar{\mathcal{M}}$ is a critical point of $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$
- $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ is strictly concave in $\bar{\mathcal{M}}$
- $\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) \rightarrow -\infty$ as $\bar{\mathcal{M}}$ approaches the boundary of the parameter space

The first of these stages is presented below, since it involves the derivation of the reestimation formulae. The remaining two assertions are demonstrated in appendices A and B.

From equation (30),

$$\frac{\partial \hat{\bar{c}}_{x,i}}{\partial \bar{\mu}_i} = \frac{\bar{\tau}_i}{K_{x,i}} \tag{31}$$

$$\frac{\partial \hat{\bar{c}}_{x,i}}{\partial \bar{\tau}_i} = \frac{\bar{\gamma}_i(d_i\bar{\mu}_i - O)}{K^2_{x,i}} \tag{32}$$

$$\frac{\partial \hat{\bar{c}}_{x,i}}{\partial \bar{\gamma}_i} = \frac{\bar{\tau}_i(O - d_i\bar{\mu}_i)}{K^2_{x,i}} \tag{33}$$

where $K_{x,i} = (\bar{\tau}_i + d_i\bar{\gamma}_i)$ and $O = \sum_{t=t_{i-1}+1}^{t_i} y_t$

From equation (12),

$$\begin{aligned} log\hat{P}(y,x|\mathcal{M}) &= log(\prod_{i=1}^{N} a_{i-1,i}\hat{P}_{\sigma_i}(y^{t_i}_{t_i-1+1})) & (34)\\ &= \sum_{i=1}^{N}(log(a_{i-1,i}) + log\hat{P}_{\sigma_i}(y^{t_i}_{t_i-1+1})) & (35)\end{aligned}$$

And from (2),

$$\begin{aligned} log\hat{P}_{\sigma_i}(y^{t_i}_{t_i-1+1}) &= logP_{\sigma_i}(y^{t_i}_{t_{i-1}+1},\hat{c}) & (36)\\ &= log\mathcal{D}_i(d_i) + log\mathcal{N}_{(\mu_i,\gamma_i)}(\hat{c}) + \sum_{t=t_{i-1}+1}^{t_i} log\mathcal{N}_{(\hat{c},\tau_i)}(y_t) & (37)\end{aligned}$$

Derivation of $\bar{\mu}_i$

From equations (26), (35) and (37),

$$\begin{aligned} \frac{\partial}{\partial\bar{\mu}_i}\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) &= \sum_x \hat{P}(y,x|\mathcal{M})\frac{\partial}{\partial\bar{\mu}_i}log\hat{P}(y,x|\bar{\mathcal{M}}) & (38)\\ &= \sum_x \hat{P}(y,x|\mathcal{M})\sum_{j=1}^{N}\frac{\partial}{\partial\bar{\mu}_i}log\hat{P}_{\sigma_j}(y^{t_j}_{t_j-1+1}) & (39)\\ &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})\frac{\partial}{\partial\bar{\mu}_i}log\hat{P}_{\sigma_i}(y^{t_i}_{t_{i-1}+1}) & (40)\\ &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})(\frac{\partial}{\partial\bar{\mu}_i}log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) & \\ &+ \sum_{t=t_{i-1}+1}^{t_i}\frac{\partial}{\partial\bar{\mu}_i}log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t)) & (41)\end{aligned}$$

Now, using (31)

$$\frac{\partial}{\partial\bar{\mu}_i}log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) = \frac{d_i(\hat{\bar{c}}_{x,i} - \bar{\mu}_i)}{K_{x,i}} \tag{42}$$

and,

$$\frac{\partial}{\partial\bar{\mu}_i} log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t) = \frac{(y_t - \hat{\bar{c}}_{x,i})}{K_{x,i}} \qquad (43)$$

Therefore,

$$\frac{\partial}{\partial\bar{\mu}_i}\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) = \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})(\frac{d_i(\hat{\bar{c}}_{x,i}-\bar{\mu}_i)}{K_{x,i}} + \sum_{t=t_{i-1}+1}^{t_i}\frac{(y_t-\hat{\bar{c}}_{x,i})}{K_{x,i}}) \qquad (44)$$

Setting the partial derivative to zero, to obtain a critical point, and multiplying through by $K_{x,i}$ gives,

$$\begin{aligned} 0 &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})(d_i(\hat{\bar{c}}_{x,i}-\bar{\mu}_i) + \sum_{t=t_{i-1}+1}^{t_i}(y_t-\hat{\bar{c}}_{x,i})) && (45) \\ &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})(\sum_{t=t_{i-1}+1}^{t_i} y_t - d_i\bar{\mu}_i) && (46) \end{aligned}$$

It follows that,

$$\bar{\mu}_i = \frac{\sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})\sum_{t=t_{i-1}+1}^{t_i} y_t}{\sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})d_i} \qquad (47)$$

which is the required result.

Derivation of $\bar{\tau}_i$

As in the derivation of $\bar{\mu}$,

$$\begin{aligned} \frac{\partial}{\partial\bar{\tau}_i}\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})(\frac{\partial}{\partial\bar{\tau}_i}log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) \\ &\quad + \sum_{t=t_{i-1}+1}^{t_i}\frac{\partial}{\partial\bar{\tau}_i}log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t)) && (48) \end{aligned}$$

But, using (32),

$$\frac{\partial}{\partial\bar{\tau}_i}log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) = \frac{(\bar{\mu}_i - \hat{\bar{c}}_{x,i})(d_i\bar{\mu}_i - O)}{K^2_{x,i}} \qquad (49)$$

Also, again using (32),

$$\frac{\partial}{\partial\bar{\tau}_i}log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t) = \frac{-1}{2\bar{\tau}_i} - \frac{1}{2\bar{\tau}_i^2}(\frac{2\bar{\tau}_i\bar{\gamma}_i(\hat{\bar{c}}_{x,i}-y_t)(d_i\bar{\mu}_i - O)}{K^2_{x,i}} - (\hat{\bar{c}}_{x,i}-y_t)^2) \qquad (50)$$

from which it follows that,

$$\sum_{t=t_{i-1}+1}^{t_i} \frac{\partial}{\partial \bar{\tau}_i} log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t) = -\frac{d_i}{2\bar{\tau}_i} - \frac{1}{2\bar{\tau}_i^2}(\frac{2\bar{\tau}_i\bar{\gamma}_i(d_i\hat{\bar{c}}_{x,i} - O)(d_i\bar{\mu}_i - O)}{K_{x,i}^2} - \sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2) \tag{51}$$

Therefore at a critical point, combining equations (48), (49) and (51),

$$\begin{aligned} 0 &= \frac{\partial}{\partial \bar{\tau}_i}\hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) \\ &= \sum_{x\in S_i} \hat{P}(y, x|\mathcal{M})[\frac{(\bar{\mu}_i - \hat{\bar{c}}_{x,i})(d_i\bar{\mu}_i - O)}{K_{x,i}^2} \\ &\quad +(-\frac{d_i}{2\bar{\tau}_i} - \frac{1}{2\bar{\tau}_i^2}(\frac{2\bar{\tau}_i\bar{\gamma}_i(d_i\hat{\bar{c}}_{x,i} - O)(d_i\bar{\mu}_i - O)}{K_{x,i}^2} \\ &\quad - \sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2))] \end{aligned} \tag{52}$$

Consider the term in square brackets. Multiplying by -1 and rearanging gives,

$$\frac{(d_i\bar{\mu}_i - O)}{K_{x,i}^2}\{-(\bar{\mu}_i - \hat{\bar{c}}_{x,i}) + \frac{\bar{\gamma}_i}{\bar{\tau}_i}(d_i\hat{\bar{c}}_{x,i}) - O)\} + \frac{d_i}{s\bar{\tau}_i} - \frac{1}{2\bar{\tau}_i^2}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2 \tag{53}$$

Multiplying by $\bar{\tau}_i$ and expanding the terms in curly brackets gives

$$\frac{(d_i\bar{\mu}_i - O)}{K_{x,i}^2}\{-\bar{\tau}_i\bar{\mu}_i + \bar{\tau}_i\hat{\bar{c}}_{x,i} + d_i\hat{\bar{c}}_{x,i}\bar{\gamma}_i - \bar{\gamma}_iO\} + \frac{d_i}{2} - \frac{1}{2\bar{\tau}_i}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2 \tag{54}$$

Now consider the term in curly brackets. This can be rewritten as,

$$\begin{aligned} -\bar{\tau}_i\bar{\mu}_i - \bar{\gamma}_iO + \hat{\bar{c}}_{x,i}(\bar{\tau}_i + d_i\bar{\gamma}_i) &= -(\bar{\tau}_i\bar{\mu}_i + \bar{\gamma}_iO) + (\bar{\tau}_i\bar{\mu}_i + \bar{\gamma}_iO) \\ &= 0 \end{aligned}$$

from the definition of $\hat{\bar{c}}_{x,i}$.

Therefore equation (52) reduces to

$$0 = \sum_{x\in S_i}\hat{P}(y, x|\mathcal{M})[\frac{d_i}{2} - \frac{1}{2\bar{\tau}_i}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2] \tag{55}$$

From which it follows that,

$$\bar{\tau}_i = \frac{\sum_{x\in S_i}\hat{P}(y, x|\mathcal{M})\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2}{\sum_{x\in S_i}\hat{P}(y, x|\mathcal{M})d_i} \tag{56}$$

Derivation of $\bar{\gamma}_i$

$$\begin{aligned}\frac{\partial}{\partial\bar{\gamma}_i}\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})(\frac{\partial}{\partial\bar{\gamma}_i}log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) \\ &+ \sum_{t=t_{i-1}+1}^{t_i}\frac{\partial}{\partial\bar{\gamma}_i}log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t)) \qquad (57)\end{aligned}$$

Using (33),

$$\frac{\partial}{\partial\bar{\gamma}_i}log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) = -\frac{1}{2\bar{\gamma}_i} - \frac{1}{2\bar{\gamma}_i^2}(-\frac{2\bar{\gamma}_i\bar{\tau}_i(\bar{\mu}_i - \hat{\bar{c}}_{x,i})}{K_{x,i}^2} - (\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2) \qquad (58)$$

Also, again using (33),

$$\frac{\partial}{\partial\bar{\gamma}_i}log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t) = -\frac{1}{K_{x,i}^2}(\hat{\bar{c}}_{x,i} - y_t)(O - \bar{\mu}_i d_i) \qquad (59)$$

Therefore,

$$\sum_{t=t_{i-1}+1}^{t_i}\frac{\partial}{\partial\bar{\gamma}_i}log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t) = -\frac{(O - \bar{\mu}_i d_i)}{K_{x,i}^2}(d_i\hat{\bar{c}}_{x,i} - O) \qquad (60)$$

Therefore at a critical point, combining equations (57), (58) and (60),

$$\begin{aligned}0 &= \frac{\partial}{\partial\bar{\gamma}_i}\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) \\ &= \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})[-\frac{1}{2\bar{\gamma}_i} \\ &\quad -\frac{1}{2\bar{\gamma}_i^2}(\frac{-2\bar{\gamma}_i\bar{\tau}_i(\bar{\mu}_i - \hat{\bar{c}}_{x,i})(O - \bar{\mu}_i d_i)}{K_{x,i}^2} - (\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2) \\ &\quad -\frac{(O - \bar{\mu}_i d_i)(d_i\hat{\bar{c}}_{x,i} - O)}{K_{x,i}^2}] \qquad (61)\end{aligned}$$

Consider the term in square brackets. Multiplying this term by $-2\bar{\gamma}_i$ and then rearranging gives,

$$\begin{aligned}1 &+ \frac{2(O - \bar{\mu}_i d_i)}{K_{x,i}^2}\{\hat{\bar{c}}_{x,i}(\bar{\tau}_i + \bar{\gamma}_i d_i) - (\bar{\tau}_i\bar{\mu}_i + \bar{\gamma}_i O)\} \\ &- \frac{1}{\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2 \\ &= 1 - \frac{1}{\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2 \qquad (62)\end{aligned}$$

because, from the definition of $\hat{\bar{c}}_{x,i}$, the term in curly brackets is equal to 0.

Therefore equation (57) reduces to,

$$0 = \sum_{x \in S_i} \hat{P}(y, x|\mathcal{M})[1 - \frac{1}{\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2] \tag{63}$$

It follows that

$$\bar{\gamma}_i = \frac{\sum_{x \in S_i} \hat{P}(y, x|\mathcal{M})(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2}{\sum_{x \in S_i} \hat{P}(y, x|\mathcal{M})} \tag{64}$$

This concludes the derivation of the reestimation formulae for parameters μ_i, τ_i and γ_i.

8.2 Remarks on the derivation of the reestimation formulae

As with the standard reestimation formula for the variance of a gaussian state in a conventional HMM, the reestimated value of the state mean $\bar{\mu}$ appears on the right hand side of equation (28). However, because $\hat{\bar{c}}_{x,i}$ is a function of $\bar{\gamma}$, the term $\bar{\gamma}$ appears on both sides of equation (28). For the purposes of implementation it is natural to use the old values μ and γ on the right-hand side of equation (28). The implications of this will be investigated experimentally. The analogous remarks hold for the quantity $\bar{\tau}$ in equation (29).

The assumptions

(i) $\bar{\gamma}_i > \bar{\tau}_i$ for all i, and

(ii) the sequence $y = y_1, ..., y_T$ is not constant

are sufficient to ensure that the critical point of the auxilliary function is a unique maximum. It is noted in section B.2 that the way in which these assumptions are used in the proof suggests that they may also be necessary.

9 Conclusions

This memorandum has presented a new segmental HMM which addresses some of the limitations of conventional HMMs in the context of speech pattern modelling. The main features of the new model are:

- The use of an underlying semi-Markov process to model speech patterns at the segment level, and
- A segment model in which separate processes are used to model extra- segment and intra-segment variability.

It has been shown that the model is computationally useful to the extent that it admits extensions of the conventional HMM classification and parameter estimation algorithms.

It has been shown that segmental HMMs can be regarded as an extension and integration of conventional variable frame rate analysis and hidden Markov modelling. In addition, the relationship between gaussian segmental HMMs and continuous gaussian mixture HMMs has been explored.

Segmental HMMs ensure that extra-segmental factors, such as choice of acoustic target or identity of speaker, are fixed throughout a segment rather than being allowed to vary in synchrony with the speech pattern feature vectors as in a conventional HMM. At present, they do not ensure that factors such as identity of speaker are preserved between segments, nor do they model the dynamic nature of speech patterns. However, it is hoped that by identifying the target pdfs more closely with parameters which reflect these factors, for example articulatory parameters, the type of segmental HMMs described here can be extended to address these issues in a similar manner to that described in [1]

References

[1] R Bakis, "Coarticulation Modelling with Continuous-State HMMs", Proc. 1991 IEEE Workshop on Automatic Speech Recognition, Arden House, Harriman, NY, December 15-18, 1991, pp 20-21.

[2] L E Baum, T Petrie, G Soules and N Weiss, "A maximisation technique occuring in the statistical analysis of probabilistic functions of Markov chains", The Annals of Mathematical Statistics, Vol. 41, No. 1, pp 164-171, 1970.

[3] H Bourlard and C J Wellekens, "Connected digit recognitionby phonemic semi-Markov chains for state occupancy modelling", Proc EUSIPCO-86.

[4] J S Bridle and N C Sedgwick, "A method for segmenting acoustic patterns with applications to automatic speech recognition", Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, pp 656-659, 1977.

[5] B-H Juang, "Maximum-likelihood estimation for mixture multivariate stochastic observations of Markov chains", AT&T Tech. J., vol 64, no. 4, pp 1235-1249, 1985.

[6] S E Levinson, "Continuously variable duration hidden Markov models for automatic speech recognition", Computer Speech and Language, Volume 1, Number 1, pp 29-46, March 1986.

[7] L Liporace, "Maximum likelihood estimation for multivariate observations of Markov sources", IEEE Trans. Information Theory, vol IT-28, 5, 1982.

[8] S M Peeling and K M Ponting, "Variable frame rate analysis in the ARM continuous speech recognition system", Speech Communication 10, pp 155-162, 1991.

[9] A G Richter, "Modeling of continuous speech observations", presented at IBM Europe Institute meeting on "Advances in Speech Processing", Oberlech, 1986.

[10] M J Russell, "Maximum likelihood hidden semi-Markov model parameter estimation for automatic speech recognition", RSRE Memorandum 3837, July 1985.

[11] M J Russell and R K Moore, "Explicit modelling of state occupancy in hidden Markov models for automatic speech recognition", Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, pp 5-8, 1985.

[12] M J Russell and A E Cook, "Experimental evaluation of duration modelling techniques for automatic speech recognition", Proc IEEE Int Conf on Acoustics, Speech and Signal Processing, pp 2376-2379, 1987.

INTENTIONALLY BLANK

A Proof of the concavity of $\hat{Q}(\mathcal{M}, \bar{\mathcal{M}})$

Claim In the notation of section 8, if $\bar{\gamma}_i > \bar{\tau}_i$ for all i, then $\hat{Q}(\mathcal{M}, \bar{\mathcal{M}})$ is strictly concave in $\bar{\mathcal{M}}$.

Proof

It is sufficient to show that

$$\frac{\partial^2}{\partial \lambda^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) < 0 \tag{65}$$

for $\lambda = \bar{\mu}_i$, $\bar{\gamma}_i$ and $\bar{\tau}_i$, for all i.

Claim: $\frac{\partial^2}{\partial \bar{\mu}^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) < 0$

This is straightforward. Differentiating equation (44) gives,

$$\begin{aligned} \frac{\partial^2}{\partial \bar{\mu}_i^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) &= \sum_{x \in S_i} \hat{P}(y, x|\mathcal{M}) (\frac{-d_i^2 \bar{\gamma}_i}{K_{x,i}} - \sum_{t=t_{i-1}+1}^{t_i} \frac{\bar{\tau}_i}{K_{x,i}^2}) \\ &< 0 \end{aligned} \tag{66}$$

since d_i, $\bar{\gamma}_i$, $\bar{\tau}_i$ and $K_{x,i}$ are all strictly positive.

Claim: $\frac{\partial^2}{\partial \bar{\gamma}^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) < 0$

From the derivation of equation (63),

$$\frac{\partial}{\partial \bar{\gamma}_i} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) = \sum_{x \in S_i} \hat{P}(y, x|\mathcal{M}) [\frac{-1}{2\bar{\gamma}_i} (1 - \frac{1}{\bar{\gamma}_i} (\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2)] \tag{67}$$

Therefore, differentiating again with respect to $\bar{\gamma}_i$,

$$\frac{\partial^2}{\partial \bar{\gamma}_i^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) = - \sum_{x \in S_i} \hat{P}(y, x|\mathcal{M}) \frac{1}{2\bar{\gamma}_i^3} (\frac{2\bar{\gamma}_i \bar{\tau}_i (\bar{\mu}_i - \hat{\bar{c}}_{x,i})(O - d_i \bar{\mu}_i)}{K_{x,i}^2} - \bar{\gamma}_i + 2(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2) \tag{68}$$

Now, it follows from the definition of $\hat{\bar{c}}_{x,i}$ that

$$(O - d_i \bar{\mu}_i) = - \frac{K_{x,i}}{\bar{\gamma}_i} (\bar{\mu}_i - \hat{\bar{c}}_{x,i}) \tag{69}$$

Therefore, from equation (68),

$$\frac{\partial^2}{\partial \bar{\gamma}_i^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) = - \sum_{x \in S_i} \hat{P}(y, x|\mathcal{M}) \frac{1}{2\bar{\gamma}_i^3} ((\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2 (2 - \frac{2\bar{\tau}_i}{K_{x,i}}) - \bar{\gamma}_i) \tag{70}$$

But, since $\bar{\gamma}_i > \bar{\tau}_i$, it follows that $\frac{2\bar{\tau}_i}{K_{x,i}} < 1$. Hence

$$\begin{aligned} \frac{\partial^2}{\partial \bar{\gamma}_i^2} \hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) &< - \sum_{x \in S_i} \hat{P}(y, x|\mathcal{M}) \frac{1}{2\bar{\gamma}_i^3} ((\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2 - \bar{\gamma}_i) && (71) \\ &= 0 && (72) \end{aligned}$$

from the definition of $\bar{\gamma}_i$ (equation (28)).

Claim: $\frac{\partial^2}{\partial \bar{\tau}^2}\hat{\mathcal{Q}}(\mathcal{M},\bar{\mathcal{M}}) < 0$

Using the derivation of equation (55), it is seen that

$$\frac{\partial}{\partial \bar{\tau}_i}\hat{\mathcal{Q}}(\mathcal{M},\bar{\mathcal{M}}) = -\sum_{x\in S_i} \hat{P}(y,x|\mathcal{M})[\frac{1}{\bar{\tau}_i}(\frac{d_i}{2} - \frac{1}{2\bar{\tau}_i}\sum_{t=t_{i-1}+1} t_i(\hat{\bar{c}}_{x,i} - y_t)^2] \tag{73}$$

It follows that,

$$\begin{aligned}\frac{\partial^2}{\partial \bar{\tau}_i^2}\hat{\mathcal{Q}}(\mathcal{M},\bar{\mathcal{M}}) &= \sum_{x\in S_i} \hat{P}(y,x|\mathcal{M})[\frac{1}{\bar{\tau}_i^2}(\frac{d_i}{2} \\ &+ (\frac{\bar{\gamma}_i}{K_{x,i}^2}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)(d_i\bar{\mu}_i - O) - \frac{1}{\bar{\tau}_i}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2))]\end{aligned} \tag{74}$$

From the definition of $\hat{\bar{c}}_{x,i}$,

$$d_i\bar{\mu}_i - O = \frac{K_{x,i}}{\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i}) \tag{75}$$

and

$$\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t) = \frac{\bar{\tau}_i}{\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i}) \tag{76}$$

Substituting the last two results into equation (74) gives

$$\begin{aligned}\frac{\partial^2}{\partial \bar{\tau}_i^2}\hat{\mathcal{Q}}(\mathcal{M},\bar{\mathcal{M}}) &= \sum_{x\in S_i} \hat{P}(y,x|\mathcal{M})[\frac{1}{\bar{\tau}_i^2}(\frac{d_i}{2} + \frac{\bar{\tau}_i}{K_{x,i}\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2 \\ &- \frac{1}{\bar{\tau}_i}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2)] &(77) \\ &< \sum_{x\in S_i} \hat{P}(y,x|\mathcal{M})[\frac{1}{\bar{\tau}_i^2}(\frac{d_i}{2} + \frac{1}{2\bar{\gamma}_i}(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2 \\ &- \frac{1}{\bar{\tau}_i}\sum_{t=t_{i-1}+1}^{t_i}(\hat{\bar{c}}_{x,i} - y_t)^2)] &(78) \\ &= \frac{1}{\bar{\tau}_i^2}[\frac{\mathcal{D}}{2} + \frac{\mathcal{I}\bar{\gamma}_i}{2\bar{\gamma}_i} - \frac{\mathcal{D}\bar{\tau}_i}{2\bar{\tau}_i}] &(79) \\ &\leq 0 &(80)\end{aligned}$$

where $\mathcal{D} = \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})d_i$ and $\mathcal{I} = \sum_{x\in S_i}\hat{P}(y,x|\mathcal{M})$ Inequality (78) holds because it is assumed that $\bar{\gamma}_i > \bar{\tau}_i$. Equation (79) follows from the definitions of $\bar{\gamma}_i$ (28) and $\bar{\tau}_i$ (29). The final inequality follows from the fact that $\mathcal{D} \geq \mathcal{I}$.

B $\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) \rightarrow -\infty$ as $\bar{\mathcal{M}}$ approaches the boundary of the parameter space

This section uses the notation of the main body of the memorandum. The assumptions that $\bar{\gamma}_i > \bar{\tau}_i$ and that the observation sequence $y = y_1, \ldots, y_T$ is not constant are both used in the proof.

B.1 Proof

Focussing again on the parameters $\bar{\mu}_i$, $\bar{\gamma}_i$ and $\bar{\tau}_i$, the cases which must be considered are:

- $\bar{\mu}_i \rightarrow \pm\infty$
- $\bar{\gamma}_i \rightarrow 0$ or $\bar{\gamma}_i \rightarrow \infty$
- $\bar{\tau}_i \rightarrow 0$ or $\bar{\tau}_i \rightarrow \infty$

From equation (26),

$$\begin{aligned}
\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) &= \sum_x \hat{P}(y,x|\mathcal{M}) log \hat{P}(y,x|\bar{\mathcal{M}}) \\
&= \sum_x \hat{P}(y,x|\mathcal{M}) \sum_{i=1}^{N} (log(a_{i-1,i}) + log(\mathcal{D}_i(d_i)) \\
&+ \; log \mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) + \sum_{t=t_{i-1}+1}^{t_i} log \mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t)) \qquad (81)
\end{aligned}$$

Hence it is sufficient to show that

$$log \mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) = -\frac{1}{2} log(2\pi) - \frac{1}{2} log(\bar{\gamma}_i) - \frac{(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2}{2\bar{\gamma}_i} \qquad (82)$$

and

$$\sum_{t=t_{i-1}+1}^{t_i} log \mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t) = \sum_{t=t_{i-1}+1}^{t_i} (-\frac{1}{2} log(2\pi) - \frac{1}{2} log(\bar{\tau}_i) - \frac{(\hat{\bar{c}}_{x,i} - y_t)^2}{2\bar{\tau}_i}) \qquad (83)$$

tend to $-\infty$ in all of the cases listed above.

<u>Case 1: $\bar{\mu}_i \rightarrow \pm\infty$</u>

This is straightforward. To see that (82) tends to $-\infty$ note that

$$\begin{aligned}
\frac{(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2}{2\bar{\gamma}_i} &= \frac{(\bar{\mu}_i d_i - O)^2 \bar{\gamma}_i}{2(\bar{\tau}_i + d_i \bar{\gamma}_i)^2} \qquad (84) \\
&\rightarrow \infty \text{ as } \bar{\mu}_i \rightarrow \pm\infty \qquad (85)
\end{aligned}$$

and, in the case of equation (83),

$$\begin{aligned} \frac{(\hat{\bar{c}}_{x,i} - y_t)^2}{2\bar{\tau}_i} &= \frac{(\bar{\tau}_i\bar{\mu}_i + O\bar{\gamma}_i - y_t(\bar{\tau}_i + d_i\bar{\gamma}_i)^2)}{2(\bar{\tau}_i + d_i\bar{\gamma}_i)^2\bar{\tau}_i} && (86) \\ &\rightarrow \infty \text{ as } \bar{\mu}_i \rightarrow \pm\infty && (87) \end{aligned}$$

Hence, since all other terms in (82) and (83) are independent of $\bar{\mu}_i$, $\hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) \rightarrow -\infty$ as $\bar{\mu}_i \rightarrow \pm\infty$

Case 2.1: $\bar{\gamma}_i \rightarrow 0$

As above

$$\begin{aligned} \frac{(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2}{2\bar{\gamma}_i} &= \frac{(\bar{\mu}_i d_i - O)^2\bar{\gamma}_i}{2(\bar{\tau}_i + d_i\bar{\gamma}_i)^2} && (88) \\ &> \frac{(\bar{\mu}_i d_i - O)^2}{2(d_i + 1)^2\bar{\gamma}_i} && (89) \end{aligned}$$

because $\bar{\gamma}_i > \bar{\tau}_i$. Hence $log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)} \rightarrow -\infty$ as $\bar{\gamma}_i \rightarrow 0$, because $\frac{1}{\bar{\gamma}_i} \rightarrow \infty$ faster than $log(\bar{\gamma}_i) \rightarrow -\infty$ as $\bar{\gamma}_i \rightarrow 0$. Similarly, from (86),

$$\begin{aligned} \frac{(\hat{\bar{c}}_{x,i} - y_t)^2}{2\bar{\tau}_i} &= \frac{(\bar{\tau}_i\bar{\mu}_i + O\bar{\gamma}_i - y_t(\bar{\tau}_i + d_i\bar{\gamma}_i)^2)}{2(\bar{\tau}_i + d_i\bar{\gamma}_i)^2\bar{\tau}_i} && (90) \\ &> \frac{(\bar{\mu}_i\bar{\tau}_i + \bar{\gamma}_i - y_t(\bar{\tau}_i + d_i\bar{\gamma}_i))^2}{2(d_i + 1)^2\bar{\gamma}_i^3} && (91) \\ &\rightarrow \infty \text{ as } \bar{\gamma}_i \rightarrow 0 && (92) \end{aligned}$$

Hence $\hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) \rightarrow -\infty$ as $\bar{\gamma}_i \rightarrow 0$

Case 2.2: $\bar{\gamma}_i \rightarrow \infty$

From equation (84),

$$\begin{aligned} \frac{(\bar{\mu}_i - \hat{\bar{c}}_{x,i})^2}{2\bar{\gamma}_i} &\doteq \frac{(\bar{\mu}_i d_i - O)^2\bar{\gamma}_i}{2(\bar{\tau}_i + d_i\bar{\gamma}_i)^2} \\ &\rightarrow 0 \text{ as } \bar{\gamma}_i \rightarrow \infty && (93) \end{aligned}$$

Therefore

$$log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) \rightarrow -\infty \text{ as } \bar{\gamma}_i \rightarrow \infty \qquad (94)$$

Also, since

$$\begin{aligned} \hat{\bar{c}}_{x,i} &= \frac{\bar{\mu}_i\bar{\tau}_i + O\bar{\gamma}_i}{\bar{\tau}_i + d_i\bar{\gamma}_i} \\ &\rightarrow \frac{O}{d_i} \text{ as } \bar{\gamma}_i \rightarrow \infty && (95) \end{aligned}$$

it follows that $log\mathcal{N}_{(\hat{\bar{c}}_{x,i},\bar{\tau}_i)}(y_t)$ is bounded as $\bar{\gamma}_i \rightarrow \infty$.

Hence $\hat{Q}(\mathcal{M}, \bar{\mathcal{M}}) \rightarrow -\infty$ as $\bar{\gamma}_i \rightarrow \infty$, as required.

Case 3.1: $\bar{\tau}_i \to 0$

Since

$$\hat{\bar{c}}_{(x,i)} \to \frac{O}{d_i} \text{ as } \bar{\tau}_i \to 0 \tag{96}$$

it follows that $log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i})$ is bounded as $\bar{\tau}_i \to 0$. This leaves the term $log\mathcal{N}_{(\hat{\bar{c}}_{(x,i)},\bar{\tau}_i)}(y_t)$. The relevant contribution of this factor to $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ consists of weighted sums of the terms $-log(\bar{\tau}_i)$ and $\frac{(\hat{\bar{c}}_{x,i},-y_t)^2}{2\bar{\tau}_i}$. But

$$(\hat{\bar{c}}_{x,i} - y_t)^2 \to (\frac{O}{d_i} - y_t)^2 \text{ as } \bar{\tau}_i \to 0 \tag{97}$$

Hence, provided that $\frac{O}{d_i} \neq y_t$ for some t, $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ will tend to $-\infty$ as $\bar{\tau}_i \to 0$ This acounts for the assumption that the observation sequence $y = y_1, ..., y_T$ is not constant.

Case 3.2: $\bar{\tau}_i \to \infty$

As $\bar{\tau}_i \to \infty$, $\bar{\gamma}_i \to \infty$, since $\bar{\gamma}_i > \bar{\tau}_i$ by assumption. The argument for case 2.2 above then shows that $log\mathcal{N}_{(\bar{\mu}_i,\bar{\gamma}_i)}(\hat{\bar{c}}_{x,i}) \to -\infty$ as $\bar{\tau}_i \to \infty$

Finally,

$$\frac{(\hat{\bar{c}}_{(x,i)} - y_t)^2}{2\bar{\tau}_i} = \frac{(\bar{\tau}_i\bar{\mu}_i + O\bar{\gamma}_i - y_t(\bar{\tau}_i + d_i\bar{\gamma}_i))^2}{2\bar{\tau}_i(\bar{\tau}_i + d_i\bar{\gamma}_i)^2} \tag{98}$$

$$\to 0 \text{ as } \bar{\tau}_i \to \infty \tag{99}$$

because $\bar{\gamma}_i > \bar{\tau}_i$. Hence $log\mathcal{N}_{(\hat{\bar{c}}_{(x,i)},\bar{\tau}_i)} \to -\infty$ as $\bar{\tau}_i \to \infty$. It follows that $\hat{Q}(\mathcal{M},\bar{\mathcal{M}}) \to -\infty$ as $\bar{\tau}_i \to \infty$ as claimed.

B.2 Remarks

The above arguments suggest that the condition $\bar{\gamma}_i > \bar{\tau}_i$ may be necessary as well as sufficient. In case 2.1, if it is not the case that $\bar{\tau}_i \to 0$ as $\bar{\gamma}_i \to 0$, then it is possible for the terms in eqautions (88) and (90) to be b unded above as $\bar{\tau}_i \to 0$. In this case the term $-log(\bar{\gamma}_i)$ will dominate and $\hat{Q}(\mathcal{M},\bar{\mathcal{M}})$ will tend to infinity as $\bar{\gamma}_i \to 0$. Consequently the point identified by the reestimation formulae will no longer necessarily be a maximum of the auxiliary function.

INTENTIONALLY BLANK

REPORT DOCUMENTATION PAGE

DRIC Reference Number (if known) ....

Overall security classification of sheet ..........UNCLASSIFIED..........

(As far as possible this sheet should contain only unclassified information. If it is necessary to enter classified information, the field concerned must be marked to indicate the classification, eg (R), (C) or (S).

Originators Reference/Report No. MEMO 4599	Month JULY	Year 1992

Originators Name and Location
DRA, ST ANDREWS ROAD
MALVERN, WORCS WR14 3PS

Monitoring Agency Name and Location

Title

A SEGMENTED HIDDEN MARKOV MODEL FOR SPEECH PATTERN PROCESSING

Report Security Classification UNCLASSIFIED	Title Classification (U, R, C or S) U

Foreign Language Title (in the case of translations)

Conference Details

Agency Reference	Contract Number and Period
Project Number	Other References

Authors RUSSELL, M J	Pagination and Ref 32

Abstract

A simple statistical segmental approach to speech pattern modelling, based on segmental hidden Markov models, is proposed which addresses some of the limitations of conventional hidden Markov model based methods. The most important features of the new approach are the use of an underlying semi-Markov process to model speech at the segment level, rather than time-synchronous frame level, and to enable improved segment duration modelling, and the development of a segment model in which separate statistical processes are used to characterise extra-state and intra-state variability, thus making the temporal independence assumption more acceptable within a segment. A basic mathematical analysis of gaussian segmental hidden Markov models is presented and model parameter re-estimation equations are derived. The relationship between the new type of model and variable frame rate analysis and conventional gaussian mixture based hidden Markov models is exposed.

Abstract Classification (U, R, C or S)
U

Descriptors

Distribution Statement (Enter any limitations on the distribution of the document)
UNLIMITED

INTENTIONALLY BLANK